DNA SEQUENCES FOR AN AMINO ACID TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS CONTAINING A TRANSPORTER AND THEIR USE

Ins. A1

BACKGROUND OF THE INVENTION

1. Field of the Invention:

The present invention relates to DNA sequences that contain the coding region of amino acid transporters, whose introduction in a plant genome modifies the transfer of metabolites in transgenic plants, plasmids, bacteria, yeasts and plants containing these DNA sequences, as well as their use.

For many plant species it is known that the delivery of energy-rich compounds to the phloem through the cell wall takes place throughout the cell. Transporter molecules which allow the penetration of amino acids through the plant cell wall are not known.

In bacteria, numerous amino acid transport systems have been characterized. For aromatic amino acids, 5 different transporters have been described which can transport any one of phenylalanine, tyrosine and tryptophan, while the other transporters are specific for individual amino acids (see Sarsero et al., 1991, J Bacteriol 173: 3231-3234). The speed constants of the transport process indicates that the specific transport is less efficient. For several transporter proteins, the corresponding genes have been cloned. This has been achieved using transport-deficient mutants which were selected for their transport ability after transformation

with DNA fragments as inserts in expression vectors (see Wallace et al., 1990, J Bacteriol 172: 3214-3220). The mutants were selected depending on their ability to grow in the presence of toxic analogues of amino acids, since the mutants cannot take these up and therefore cannot be impaired.

Corresponding complementation studies have been carried out with the eukaryotic yeast, *Saccharomyces cerevisiae*. Tanaka & Fink (1985, Gene 38: 205-214) describe a histidine transporter that was identified by complementation of a mutation. Vandenbol et al. (1989, Gene 83: 153-159) describe a proline transporter for *Saccharomyces cerevisiae*. The yeast possesses two different permeases for proline. One transports with lower efficiency and can be used also for other amino acids, and the other is proline-specific and works with high affinity. The latter was coded from the *put4* gene. This carries an open reading frame for a peptide with a molecular weight of 69 kDa. The protein contains 12 membrane-penetrating regions, but does not contain any N-terminal signal sequence for secretion. This is a typical property of integral membrane proteins. The permeases process homology for arginine and for histidine permease from yeast, but not, however, for proline permease from *Escherichia coli*.

For plant cells, based on studies on tobacco suspension cultures, it has been found that the transport of arginine, asparagine, phenylalanine and histidine are pH and energy dependent. Since a 1,000-fold excess of leucine inhibits the transport of the other amino acids, it can be assumed, therefore, that all amino acids use

the same transporter (McDaniel et al., 1982, Plant Physio 69: 246-249). Li and Bush (1991, Plant Physiol 96: 1338-1344) determined, for aliphatic, neutral amino acids, two transport systems in plasma membrane vesicles from *Beta vulgaris*. On the one hand, alanine, methionine, glutamine and leucine displace each other on the transporter protein. On the other hand, isoleucine, valine and threonine have mutually competitive effects. In combined competition kinetic studies (Li & Bush, 1990, Plant Physiol 94: 268-277) four different transport systems have been distinguished. Besides a transporter for all neutral amino acids, which work with low affinity, there exists a high affinity type which, however, possesses low affinity for isoleucine, threonine, valine and proline. Further transporters exist for acids as well as for basic amino acids.

The transporter molecule or gene for plant transporter proteins is not known.

SUMMARY OF THE INVENTION

There are now described DNA sequences which contain the coding region of a plant amino acid transporter, and whose information contained in the nucleotide sequence allows, by integration in a plant genome, the formation of RNA, by which a new amino acid transport activity can be introduced in the plant cells or an endogenous amino acid transporter activity can be expressed.

Under the term amino transporter is to be understood, for example a cDNA sequence that codes an amino transporter from *Arabidopsis thaliana*.

The identification of the coding region of the amino acid transporter is carried out by a process which allows the isolation of plant DNA sequences which code transporter molecules by means of expression in specific mutants of yeast *Saccharomyces cerevisiae*. For this, suitable yeast mutants have to be provided which cannot take up a substance for which the coding region of the transporter molecule has to be isolated from a plant gene library.

A mutant which cannot grow in media, with proline or citrulline as the only nitrogen source, is described by Jauniaux et al. (1987), Eur J Biochem 164: 601-606).

For the preparation of yeast strains that can be used to identify plant amino acid transporters, a yeast mutant which is not able to grow in media with proline and/or citrulline as the only nitrogen source is, for example, transformed with pFL 61 plasmid, which carries, as an insert, cDNA fragments from a cDNA library from *Arabidopsis thaliana*.

Further, a double mutant JT16 (Tanaka & Fink, 1985, Gene 38: 205-214) which has a deficiency in histidine synthesis (*his4*) and in histidine uptake (*hip1*) is transformed with the described pFL 61 plasmid and cultivated in a medium with addition of histidine.

It has now surprisingly been found that, in the transformation of yeast cells, certain plant cDNA fragments can complement the yeast mutation. By analysis of the properties of the proteins coded from the cDNA it can be shown that a coding region that codes a plant amino acid transporter with a wide specificity spectrum

is responsible for the complementing of the mutation (see example 3).

Such a coding region of an amino acid transporter is shown, for example, by one of the following nucleotide sequences:

1. Sequence (Seq. ID No. 1):

```
CTTAAAACAT TTATTTTATC TTCTTCTTGT TCTCTCTTTC TCTTTCTCTC ATCACT   56

ATG AAG AGT TTC AAC ACA GAA GGA CAC AAC CAC TCC ACG GCG GAA    101
Met Lys Ser Phe Asn Thr Glu Gly His Asn His Ser Thr Ala Glu
1               5                  10                  15

TCC GGC GAT GCC TAC ACC GTG TCG GAC CCG ACA AAG AAC GTC GAT    146
Ser Gly Asp Ala Tyr Thr Val Ser Asp Pro Thr Lys Asn Val Asp
                20                  25                  30

GAA GAT GGT CGA GAG AAG CGT ACC GGG ACG TGG CTT ACG GCG AGT    191
Glu Asp Gly Arg Glu Lys Arg Thr Gly Thr Trp Leu Thr Ala Ser
                35                  40                  45

GCG CAT ATT ATC ACG GCG GTG ATA GGC TCC GGA GTG TTG TCT TTA    236
Ala His Ile Ile Thr Ala Val Ile Gly Ser Gly Val Leu Ser Leu
                50                  55                  60

GCA TGG GCT ATA GCT CAG CTT GGT TGG ATC GCA GGG ACA TCG ATC    281
Ala Trp Ala Ile Ala Gln Leu Gly Trp Ile Ala Gly Thr Ser Ile
                65                  70                  75
```

```
TTA CTC ATT TTC TCG TTC ATT ACT TAC TTC ACC TCC ACC ATG CTT   326
Leu Leu Ile Phe Ser Phe Ile Thr Tyr Phe Thr Ser Thr Met Leu
                80                  85                  90

GCC GAT TGC TAC CGT GCG CCG GAT CCC GTC ACC GGA AAA CGG AAT   371
Ala Asp Cys Tyr Arg Ala Pro Asp Pro Val Thr Gly Lys Arg Asn
                95                 100                 105

TAC ACT TAC ATG GAC GTT GTT CGA TCT TAC CTC GGT GGT AGG AAA   416
Tyr Thr Tyr Met Asp Val Val Arg Ser Tyr Leu Gly Gly Arg Lys
               110                 115                 120

GTG CAG CTC TGT GGA GTG GCA CAA TAT GGG AAT CTG ATT GGG GTC   461
Val Gln Leu Cys Gly Val Ala Gln Tyr Gly Asn Leu Ile Gly Val
               125                 130                 135

ACT GTT GGT TAC ACC ATC ACT GCT TCT ATT AGT TTG GTA GCG GTA   506
Thr Val Gly Tyr Thr Ile Thr Ala Ser Ile Ser Leu Val Ala Val
               140                 145                 150

GGG AAA TCG AAC TGC TTC CAC GAT AAA GGG CAC ACT GCG GAT TGT   551
Gly Lys Ser Asn Cys Phe His Asp Lys Gly His Thr Ala Asp Cys
               155                 160                 165

ACT ATA TCG AAT TAT CCG TAT ATG GCG GTT TTT GGT ATC ATT CAA   596
Thr Ile Ser Asn Tyr Pro Tyr Met Ala Val Phe Gly Ile Ile Gln
               170                 175                 180

GTT ATT CTT AGC CAG ATC CCA AAT TTC CAC AAG CTC TCT TTT CTT   641
Val Ile Leu Ser Gln Ile Pro Asn Phe His Lys Leu Ser Phe Leu
               185                 190                 195
```

```
TCC ATT ATG GCC GCA GTC ATG TCC TTT ACT TAT GCA ACT ATT GGA   686
Ser Ile Met Ala Ala Val Met Ser Phe Thr Tyr Ala Thr Ile Gly
                200                 205                 210

ATC GGT CTA GCC ATC GCA ACC GTC GCA GGT GGG AAA GTG GGT AAG   731
Ile Gly Leu Ala Ile Ala Thr Val Ala Gly Gly Lys Val Gly Lys
                215                 220                 225

ACG AGT ATG ACG GGC ACA GCG GTT GGA GTA GAT GTA ACC GCA GCT   776
Thr Ser Met Thr Gly Thr Ala Val Gly Val Asp Val Thr Ala Ala
                230                 235                 240

CAA AAG ATA TGG AGA TCG TTT CAA GCG GTT GGG GAC ATA GCG TTC   821
Gln Lys Ile Trp Arg Ser Phe Gln Ala Val Gly Asp Ile Ala Phe
                245                 250                 255

GCC TAT GCT TAT GCC ACG GTT CTC ATC GAG ATT CAG GAT ACA CTA   866
Ala Tyr Ala Tyr Ala Thr Val Leu Ile Glu Ile Gln Asp Thr Leu
                260                 265                 270

AGA TCT AGC CCA GCT GAG AAC AAA GCC ATG AAA AGA GCA AGT CTT   911
Arg Ser Ser Pro Ala Glu Asn Lys Ala Met Lys Arg Ala Ser Leu
                275                 280                 285

GTG GGA GTA TCA ACC ACC ACT TTT TTC TAC ATC TTA TGT GGA TGC   956
Val Gly Val Ser Thr Thr Thr Phe Phe Tyr Ile Leu Cys Gly Cys
                290                 295                 300

ATC GGC TAT GCT GCA TTT GGA AAC AAT GCC CCT GGA GAT TTC CTC  1001
Ile Gly Tyr Ala Ala Phe Gly Asn Asn Ala Pro Gly Asp Phe Leu
                305                 310                 315
```

```
ACA GAT TTC GGG TTT TTC GAG CCC TTT TGG CTC ATT GAC TTT GCA  1046
Thr Asp Phe Gly Phe Phe Glu Pro Phe Trp Leu Ile Asp Phe Ala
                320                 325                 330

AAC GCT TGC ATC GCT GTC CAC CTT ATT GGT GCC TAT CAG GTG TTC  1091
Asn Ala Cys Ile Ala Val His Leu Ile Gly Ala Tyr Gln Val Phe
                335                 340                 345

GCG CAG CCG ATA TTC CAG TTT GTT GAG AAA AAA TGC AAC AGA AAC  1136
Ala Gln Pro Ile Phe Gln Phe Val Glu Lys Lys Cys Asn Arg Asn
                350                 355                 360

TAT CCA GAC AAC AAG TTC ATC ACT TCT GAA TAT TCA GTA AAC GTA  1181
Tyr Pro Asp Asn Lys Phe Ile Thr Ser Glu Tyr Ser Val Asn Val
                365                 370                 375

CCT TTC CTT GGA AAA TTC AAC ATT AGC CTC TTC AGA TTG GTG TGG  1226
Pro Phe Leu Gly Lys Phe Asn Ile Ser Leu Phe Arg Leu Val Trp
                380                 385                 390

AGG ACA GCT TAT GTG GTT ATA ACC ACT GTT GTA GCT ATG ATA TTC  1271
Arg Thr Ala Tyr Val Val Ile Thr Thr Val Val Ala Met Ile Phe
                395                 400                 405

CCT TTC TTC AAC GCG ATC TTA GGT CTT ATC GGA GCA GCT TCC TTC  1316
Pro Phe Phe Asn Ala Ile Leu Gly Leu Ile Gly Ala Ala Ser Phe
                410                 415                 420

TGG CCT TTA ACG GTT TAT TTC CCT GTG GAG ATG CAC ATT GCA CAA  1361
Trp Pro Leu Thr Val Tyr Phe Pro Val Glu Met His Ile Ala Gln
                425                 430                 435
```

```
ACC AAG ATT AAG AAG TAC TCT GCT AGA TGG ATT GCG CTG AAA ACG  1406
Thr Lys Ile Lys Lys Tyr Ser Ala Arg Trp Ile Ala Leu Lys Thr
                440                 445                 450

ATG TGC TAT GTT TGC TTG ATC GTC TCG CTC TTA GCT GCA GCC GGA  1451
Met Cys Tyr Val Cys Leu Ile Val Ser Leu Leu Ala Ala Ala Gly
                455                 460                 465

TCC ATC GCA GGA CTT ATA AGT AGT GTC AAA ACC TAC AAG CCC TTC  1496
Ser Ile Ala Gly Leu Ile Ser Ser Val Lys Thr Tyr Lys Pro Phe
                470                 475                 480

CGG ACT ATG CAT GAG TGAGTTTGAG ATCCTCAAGA GAGTCAAAAA        1541
Arg Thr Met His Glu
                485

TATATGTAGT AGTTTGGTCT TTCTGTTAAA CTATCTGGTG TCTAAATCCA      1591

ATGAGAATGC TTTATTGCTA AAACTTCATG AATCTCTCTG TATCTACATC      1641

TTTCAATCTA ATACATATGA GCTCTTCCAA AAAAAAAAAA AAAA            1685
```

2. Sequence (Seq. ID No. 3):

```
                    CTATTTTAT AATTCCTCTT CTTTTTTGTTC          29

ATAGCTTTGT AATTATAGTC TTATTTCTCT TTAAGGCTCA ATAAGAGGAG       79

ATG GGT GAA ACC GCT GCC GCC AAT AAC CAC CGT CAC CAC CAC CAT  124
Met Gly Glu Thr Ala Ala Ala Asn Asn His Arg His His His His
1               5                  10                  15
```

```
CAC GGC CAC CAG GTC TTT GAC GTG GCC AGC CAC GAT TTC GTC CCT   169
His Gly His Gln Val Phe Asp Val Ala Ser His Asp Phe Val Pro
                20                  25                  30

CCA CAA CCG GCT TTT AAA TGC TTC GAT GAT GAT GGC CGC CTC AAA   214
Pro Gln Pro Ala Phe Lys Cys Phe Asp Asp Asp Gly Arg Leu Lys
                35                  40                  45

AGA ACT GGG ACT GTT TGG ACC GCG AGC GCT CAT ATA ATA ACT GCG   259
Arg Thr Gly Thr Val Trp Thr Ala Ser Ala His Ile Ile Thr Ala
                50                  55                  60

GTT ATC GGA TCC GGC GTT TTG TCA TTG GCG TGG GCG ATT GCA CAG   304
Val Ile Gly Ser Gly Val Leu Ser Leu Ala Trp Ala Ile Ala Gln
                65                  70                  75

CTC GGA TGG ATC GCT GGC CCT GCT GTG ATG CTA TTG TTC TCT CTT   349
Leu Gly Trp Ile Ala Gly Pro Ala Val Met Leu Leu Phe Ser Leu
                80                  85                  90

GTT ACT CTT TAC TCC TCC ACA CTT CTT AGC GAC TGC TAC AGA ACC   394
Val Thr Leu Tyr Ser Ser Thr Leu Leu Ser Asp Cys Tyr Arg Thr
                95                  100                 105

GGC GAT GCA GTG TCT GGC AAC AGA AAC TAC ACT TAC ATG GAT GCC   439
Gly Asp Ala Val Ser Gly Lys Arg Asn Tyr Thr Tyr Met Asp Ala
                110                 115                 120

GTT CGA TCA ATT CTC GGT GGG TTC AAG TTC AAG ATT TGT GGG TTG   484
Val Arg Ser Ile Leu Gly Gly Phe Lys Phe Lys Ile Cys Gly Leu
                125                 130                 135
```

```
ATT CAA TAC TTG AAT CTC TTT GGT ATC GCA ATT GGA TAC ACG ATA   529
Ile Gln Tyr Leu Asn Leu Phe Gly Ile Ala Ile Gly Tyr Thr Ile
                140                 145                 150

GCA GCT TCC ATA AGC ATG ATG GCG ATC AAG AGA TCC AAC TGC TTC   574
Ala Ala Ser Ile Ser Met Met Ala Ile Lys Arg Ser Asn Cys Phe
                155                 160                 165

CAC AAG AGT GGA GGA AAA GAC CCA TGT CAC ATG TCC AGT AAT CCT   619
His Lys Ser Gly Gly Lys Asp Pro Cys His Met Ser Ser Asn Pro
                170                 175                 180

TAC ATG ATC GTA TTT GGT GTG GCA GAG ATC TTG CTC TCT CAG GTT   664
Tyr Met Ile Val Phe Gly Val Ala Glu Ile Leu Leu Ser Gln Val
                185                 190                 195

CCT GAT TTC GAT CAG ATT TGG TGG ATC TCC ATT GTT GCA GCT GTT   709
Pro Asp Phe Asp Gln Ile Trp Trp Ile Ser Ile Val Ala Ala Val
                200                 205                 210

ATG TCC TTC ACT TAC TCT GCC ATT GGT CTA GCT CTT GGA ATC GTT   754
Met Ser Phe Thr Tyr Ser Ala Ile Gly Leu Ala Leu Gly Ile Val
                215                 220                 225

CAA GTT GCA GCG AAT GGA GTT TTC AAA GGA AGT CTC ACT GGA ATA   799
Gln Val Ala Ala Asn Gly Val Phe Lys Gly Ser Leu Thr Gly Ile
                230                 235                 240

AGC ATC GGA ACA GTG ACT CAA ACA CAG AAG ATA TGG AGA ACC TTC   844
Ser Ile Gly Thr Val Thr Gln Thr Gln Lys Ile Trp Arg Thr Phe
                245                 250                 255
```

```
CAA GCA CTT GGA GAC ATT GCC TTT GCG TAC TCA TAC TCT GTT GTC   889
Gln Ala Leu Gly Asp Ile Ala Phe Ala Tyr Ser Tyr Ser Val Val
                260                 265                 270

CTA ATC GAG ATT CAG GAT ACT GTA AGA TCC CCA CCG GCG GAA TCG   934
Leu Ile Glu Ile Gln Asp Thr Val Arg Ser Pro Pro Ala Glu Ser
                275                 280                 285

AAA ACG ATG AAG AAA GCA ACA AAA ATC AGT ATT GCC GTC ACA ACT   979
Lys Thr Met Lys Lys Ala Thr Lys Ile Ser Ile Ala Val Thr Thr
                290                 295                 300

ATC TTC TAC ATG CTA TGT GGC TCA ATG GGT TAT GCC GCT TTT GGA  1024
Ile Phe Tyr Met Leu Cys Gly Ser Met Gly Tyr Ala Ala Phe Gly
                305                 310                 315

GAT GCA GCA CCG GGA AAC CTC CTC ACC GGT TTT GGA TTC TAC AAC  1069
Asp Ala Ala Pro Gly Asn Leu Leu Thr Gly Phe Gly Phe Tyr Asn
                320                 325                 330

CCG TTT TGG CTC CTT GAC ATA GCT AAC GCC GCC ATT GTT GTC CAC  1114
Pro Phe Trp Leu Leu Asp Ile Ala Asn Ala Ala Ile Val Val His
                335                 340                 345

CTC GTT GGA GCT TAC CAA GTC TTT GCT CAG CCC ATC TTT GCC TTT  1159
Leu Val Gly Ala Tyr Gln Val Phe Ala Gln Pro Ile Phe Ala Phe
                350                 355                 360

ATT GAA AAA TCA GTC GCA GAG AGA TAT CCA GAC AAT GAC TTC CTC  1204
Ile Glu Lys Ser Val Ala Glu Arg Tyr Pro Asp Asn Asp Phe Leu
                365                 370                 375
```

```
AGC AAG GAA TTT GAA ATC AGA ATC CCC GGA TTT AAG TCT CCT TAC  1249
Ser Lys Glu Phe Glu Ile Arg Ile Pro Gly Phe Lys Ser Pro Tyr
380                 385                 390

AAA GTA AAC GTT TTC AGG ATG GTT TAC AGG AGT GGC TTT GTC GTT  1294
Lys Val Asn Val Phe Arg Met Val Tyr Arg Ser Gly Phe Val Val
395                 400                 405

ACA ACC ACC GTG ATA TCG ATG CTG ATG CCG TTT TTT AAC GAC GTG  1339
Thr Thr Thr Val Ile Ser Met Leu Met Pro Phe Phe Asn Asp Val
410                 415                 420

GTC GGG ATC TTA GGG GCG TTA GGG TTT TGG CCC TTG ACG GTT TAT  1384
Val Gly Ile Leu Gly Ala Leu Gly Phe Trp Pro Leu Thr Val Tyr
425                 430                 435

TTT CCG GTG GAG ATG TAT ATT AAG CAG AGG AAG GTT GAG AAA TGG  1429
Phe Pro Val Glu Met Tyr Ile Lys Gln Arg Lys Val Glu Lys Trp
440                 445                 450

AGC ACG AGA TGG GTG TGT TTA CAG ATG CTT AGT GTT GCT TGT CTT  1474
Ser Thr Arg Trp Val Cys Leu Gln Met Leu Ser Val Ala Cys Leu
455                 460                 465

GTG ATC TCG GTG GTC GCC GGG GTT GGA TCA ATC GCC GGA GTG ATG  1519
Val Ile Ser Val Val Ala Gly Val Gly Ser Ile Ala Gly Val Met
470                 475                 480

CTT GAT CTT AAG GTC TAT AAG CCA TTC AAG TCT ACA TAT          1558
Leu Asp Leu Lys Val Tyr Lys Pro Phe Lys Ser Thr Tyr
485                 490
```

TGATGATTAT GGACCATGAA CAACAGAGAG AGTTGGTGTG TAAAGTTTAC 1608

CATTTCAAAG AAAACTCCAA AAATGTGTAT ATTGTATGTT GTTCTCATTT 1658

CGTATGGTCT CATCTTTGTA ATAAAATTTA AAACTTATGT TATAAATTAT 1708

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA 1740

SUB B1

The DNA sequences of the invention identified with the help of the transformed yeast strains, e.g., sequences Seq. No. 1 and 2, can be introduced into plasmids and thereby be combined with steering elements for expression in eukaryotic cells (see Example 4). These steering elements are, on the one hand, transcription promoters, and, on the other hand, transcription terminators. Plasmids can be used to transform eukaryotic cells with the aim of expression of a translatable mRNA which makes possible the synthesis of an amino acid transporter in the cells or with the aim of expression of a non-translatable RNA, which prevents synthesis of an endogenous amino acid transporter in the cells. The expression of an RNA corresponding to the inventive sequences of plant amino acid transporters modifies the plant acid metabolism, as well as total nitrogen metabolism. The economic significance of this modification is obvious. Nitrogen is the nutrient mainly responsible for limiting growth. The viability of germ lines as well as germination capacity of seeds is directly dependent on the nitrogen content of storage tissue. The formation of high value food materials with a high protein content is dependent on a sufficient

Nitrogen supply. Nitrogen is transported essentially in the form of amino acids. An improvement in the delivery of amino acids to their harvested parts can therefore lead to an increase in yield of agricultural plants. The possibility of forcing the uptake of amino acid in individual organs allows the qualitative improvement of such organs, which, because of the demands of the utilization process, contain little nitrogen. An example is potatoes which are grown for the production of starch. Besides this, it is possible to modify the whole plant, by which the growth of individual tissues, for example, leaves, is slowed down, while the growth of the harvested parts is increased. For this, one can imagine a lengthening of the vegetative phase of crops, which leads to an increased formation of storage substances.

Processes for the genetic modification of dicotyledonous and monocotyledonous plants are already known (see for example Gasser, C.S., Fraley, R.T., 1989, Science 244: 1293-1299; Potrykus, 191, Ann Rev Plant Mol Biol Plant Physiol 42: 205-225). For expression in plants the coding sequences must be coupled with the transcriptional regulatory elements. Such elements, called promoters, are known (EP 375091).

Further, the coding regions must be provided with transcription termination signals with which they can be correctly transcribed. Such elements are also described (see Gielen et al., 1989, EMBO J 8: 23-29). The transcriptional start region can be either native and/or homologous or foreign and/or heterologous to the host plant. If desired, termination regions are interchangeable with one another. The DNA sequence of

the transcription starting and termination regions can be prepared synthetically, obtained naturally, or can be a mixture of synthetic and natural DNA constituents. For introduction of foreign genes in higher plants, a large number of cloning vectors are available that include a replication signal for *E. coli* and a marker which allows for the selection of the transformed cells. Examples of such vectors are pBR 322, pUC-Series, M13 mp-Series, pACYC 184, etc. Depending on the method of introduction of the desired gene in the plants, other DNA sequences may be suitable. Should the Ti- or Ri-plasmid be used, e.g., for the transformation of the plant cell, then at least the right boundary, often, however, both the right and left boundary of the Ti- and Ri-Plasmid T-DNA, is attached, as a flanking region, to the gene being introduced. The use of T-DNA for the transformation of plant cells has been intensively researched and is well described in EP 120 516; Hoekama, In: The Binary Plant Vector System, Offset-drukkerij Kanters B.V. Alblasserdam (1985), Chapter V; Fraley, et al., Crit. Rev. Plant Sci., 4:1-46 and An et al. (1985) EMBO J. 4: 277-287. Once the introduced DNA is integrate in the genome, it is generally stable there and remains in the offspring of the original transformed cells. It normally contains a selection marker which induces resistance in the transformed plant cells against a biocide or antibiotic such as kanamycin, G 418, bleomycin, hygromycin or phosphinotricin, etc. The individual marker employed should therefore allow the selection of transformed cells from cells which lack the introduced DNA.

For the introduction of DNA into a plant host cell, besides transformation using *Agrobacteria*, there are many other techniques available. These techniques include the fusion of protoplasts, microinjection of DNA and electroporation, as well as ballistic methods and virus infection. From the transformed plant material, whole plants can be regenerated in a suitable medium which contains antibiotics or biocides for selection. The resulting plants can then be tested for the presence of introduced DNA. No special demands are placed on the plasmids in injection and electroporation. Simple plasmids, such as, e.g., pUC-derivatives, can be used. Should whole plants be regenerated from such transformed cells, the presence of a selectable marker gene is necessary. The transformed cells grow within the plants in the usual manner (see also McCormick et al. (1986) Plant Cell Reports 5: 81-84). These plants can be grown normally and crossed with plants that possess the same transformed genes or different genes. The resulting hybrid individuals have the corresponding phenotypical properties.

The DNA sequences of the invention can also be introduced in plasmids and thereby combined with steering elements for an expression in prokaryotic cells. The formation of a translatable RNA sequence of a eukaryotic amino acid transporter from bacteria, in spite of the considerable differences in the membrane structures of prokaryotes and eukaryotes, means that prokaryotes can now use a eukaryotic amino acid transporter with specificity for certain substrates. This makes possible the production of bacterial strains which could be used

for studies of the properties of the transporter as well as its substrate.

The invention also relates to bacteria that contain the plasmids of the invention.

The DNA sequences of the invention can also be introduced in plasmids which allow mutagenesis or a sequence modification through recombination of DNA sequences in prokaryotic or eukaryotic systems. In this way, the specificity of the amino acid transporter can be modified. Thus, the specificity of the transporter can be changed.

The invention also relates to derivatives or parts of plasmids that contain the DNA sequences of the invention and which can be used for the transformation of prokaryotic and eukaryotic cells.

By using standard processes (see Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press, NY, USA), base exchanges can be carried out or natural or synthetic sequences can be added. For binding DNA fragments with one another, adaptors or linkers can be introduced on the fragments. Further, manipulations can be carried which prepare suitable restriction cleavage sites or remove the excess DNA or restriction cleavage sites. Where insertions, deletions or substitutions such as, for example, transitions and transversions are desired, *in vitro* mutagenesis, primer repair, restrictions or ligations can be used. For methods of analysis, in general, a sequence analysis, restriction analysis and other biochemical molecular biological methods can be used. After each manipulation, the DNA sequence used can

be cleaved and bound with another DNA sequence. Each plasmid sequence can be cloned in the same or different plasmids.

Derivatives or parts of the DNA sequences and plasmids of the invention can also be used for the transformation of prokaryotic and eukaryotic cells. Further, the DNA sequences of the invention can be used according to standard processes for the isolation of similar sequences on the genome of plants of various species, which also code for amino acid or other oligosaccharide transporter molecules. With these sequence constructs, for the transformation of plant cells, can be prepared which modify the transport process in transgenic plants.

In order to specify related DNA sequences, gene libraries must first be prepared which are representative of the content of genes of a plant type or for the expression of genes in a plant type. The former are genomic libraries, while the latter are cDNA libraries. From these, related sequences can be isolated using the DNA sequences of the invention as probes. Once the related gene has been identified and isolated, a determination of the sequence and an analysis of the properties of the proteins coded from this sequence is possible.

In order to understand the examples forming the basis of this invention all the processes necessary for these tests and which are known per se will first of all be listed:

1. <u>Cloning process</u>

For cloning in *E. coli*, the vector pBluescriptSK (Short et al., 1988, Nucl Acids Res 16: 7583-7600) was used.

For the transformation of yeasts, the vector pFL61 (Minet & Lacroute, 1990, Curr Genet 18: 287-291) was used.

For the plant transformation the gene constructs in the binary vector pBIN-Hyg were cloned.

2. Bacterial and yeast strains

For the pBluescriptSK vector as well as for PBinAR constructs, the *E. coli* strain XL1blue (Bullock et al., 1987, Biotechniques, 5, 376-378) was used.

As a starting strain for the expression of the cDNA library in yeast, the yeast strain 22574d (Jauniaux et al., 1987 Eur J Biochem 164: 601-606) was used.

The transformation of the plasmids in potato plants was carried out using *Agrobacterium tumefaciens* strain LBA4404 (Bevan (1984) Nucl. Acids Res 12: 8711-8720).

3. Transformation of *Agrobacterium tumefaciens*

The transfer of the DNA in *Agrobacteria* was carried out by direct transformation by the method of Höfgen & Willmitzer (1988, Nucleic Acids Res 16: 9877). The plasmid DNA of the transformed *Agrobacterium* was isolated in accordance with the method of Birnboim and Doly (1979) (Nucl Acids Res 7: 1513-1523) and was analyzed by gel electrophoresis after suitable restriction cleavage.

4. Plant transformation

Ten small leaves, wounded with a scalpel, of a sterile potato culture were placed in 10 ml of MS medium

with 2% amino acid containing 30-50 µl of an *Agrobacterium tumefaciens* overnight culture grown under selection. After 3-5 minutes of gentle shaking, the leaves were laid out on MS medium of 1.6% glucose, 2 mg/l of zeatin ribose, 0.02 mg/l of naphthylacetic acid, 0.02 mg/l of gibberellic acid, 500 mg/l of claforan, 50 mg/l of kanamycin and 0.8% bacto agar. After incubation for one week at 25°C and 3000 lux, the claforan concentration in the medium was reduced by half.

Deposits

The following plasmids and yeast strains were deposited at the Deutschen Sammlung von Mikroorganismen (DSM) in Braunschweig, Germany on 12.06.1992 (deposit number):

Plasmid pPPP1-20 (DSM 7129)
Plasmid pBinPPP1-20 (DSM 7130)

Other features and advantages of the present invention will become apparent from the following description of the invention which refers to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the plasmid pPPP1-20 which contains the sequence Seq-ID No. 1. The finely drawn line corresponds to the sequence from pBluescriptSK. The thicker line represents the cDNA insert. The cleavage positions of the inserts are shown.

Fig. 2 shows the uptake of ^{14}C-proline from the medium.

no = time period of the uptake without competitor;

proline = time period with fourfold excess of unlabeled proline;

citrulline = time period with fourfold excess of unlabeled citrulline;

GABA = time period with fourfold excess of gamma-aminobutyric acid;

time = time in seconds;

cpm = decays counted per minute.

Fig. 3 shows the plasmid pAAP2 which contains the sequence Seq-ID No. 2. The finely drawn line corresponds to the sequence from pBluescriptSK. The thicker line represents the cDNA insert. The cleavage positions of the inserts are shown.

Fig. 4 shows a competition experiment with the yeast line 22574d::AAP2. In this experiment, the uptake of ^{14}C-labeled L-proline from the medium in the presence of a fourfold excess of other amino acids or their analogues is measured. Besides the standard abbreviations for amino acids in the three letter code, the following are also used:

Cit = citrulline;

D-Pro = D-proline;

OH-Pro = hydroxyproline; and

AC2 = azetidine-2-carboxylic acid.

Fig. 5 shows a competition experiment with the yeast line JT16::AAP2. In this experiment, the uptake of ^{14}C labeled L-histidine from the medium in the presence of a tenfold excess of other amino acids or their analogues is measured.

Besides the standard abbreviations for amino acids in the three letter code, the following are also used:

Cit = citrulline;
Orn = ornithine;
Can = canavanine; and
NH4 = ammonium.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The following examples describe the cloning and identification, as well as the function and use of a plant amino acid transporter.

Example 1

Cloning of the cDNA of a plant amino acid transporter

For complementation of the proline transport mutation of the yeast strain 22574d (Jauniaux et al., 1987, Eur J Biochem 164: 601-606) and/or the histidine synthesis and transport mutation of the strain JT16 (Tanaka & Fink 1985, Gene 38: 205-214), a cDNA of young germ lines from *Arabidopsis thaliana* (two leaf stage) in the yeast expression vector pFL61 (Minet & Lacroute), 1990 Curr Genet 18: 287-291) which had been made available by Minet (Minet et al., 1992, Plant J 2: 417-422) was used. Around 1 µg of the vector with the cDNA-insert was transformed in the yeast strain 22574d and/or JT16 by the method of Dohmen et al. (1991, Yeast 7: 691-692). Yeast transformands, which could grow in media with 4 mM proline as the sole nitrogen source or in media with 6 mM histidine, were propagated. From the lines plasmid-DNA was prepared by standard methods. Clones that could complement the particular mutation contained

plasmids with similar restriction type of the cDNA insert. These varied in size between 1.6 and 1.7 kb.

Example 2

Sequence analysis of the cDNA insert of the plasmid pFL61-ppp1-20

From a yeast line PPP1-20, obtained in a similar manner to example 1, which, in spite of the 22574d mutation, could grow with proline as the only nitrogen source, the plasmid pFL61-ppp1-20 was isolated. Its cDNA insert was prepared as a NotI fragment and cloned in the vector pBluescriptSK. In this way, the plasmid pPPP1-20 was obtained (see Figure 1). Using synthetic oligonucleotides, the insert was sequenced by the method of Sanger et al. (1977, Proc Natl Acad Sci USA 74:5463-5467). The sequence is given above (SEQ ID No. 1).

In a similar way, from a yeast line that, in spite of the *his4/hip1* double mutation, could be grown in a medium with histidine addition, the plasmid pFL61-aap2 was isolated whose insert was also cloned as a NotI fragment in pBluescriptSK. The resulting plasmid pAAP2 was sequenced and the sequence (SEQ ID No. 2) is given above. The plasmid pAAP2 has a similar structure to pPPP1-20 (see Fig. 1), but instead of the insert SEQ ID No. 1, carries the insert SEQ ID No. 2 (see Fig. 3).

Example 3

Uptake studies with ^{14}C-labeled protein into the yeast line PPP1-20 and AAP2

The yeast lines 22574d::PPP1-20 and 22574d::AAP2 that were obtained in a similar manner to Example 1 were grown in liquid medium until the culture reached the logarithmic phase. After centrifuging the culture, the cells are washed and taken up in 100 mm tris/HCl pH 4.5, 2mM $MgCl_2$ and 0.6M sorbitol. Around 100 µL of the suspension was added to a solution of 0.5mM L-proline plus 1 µCi ^{14}C labeled L-proline in 100 µL of the same buffer. The uptake of the labeled amino acid was measured by the process described by Cirillo (1989, Meth Enzymol 174: 617-622). The uptake of the labeled amino acid was compared, on the one hand, in co-incubation with protein modifying substance diethyl pyrocarbonate which is an inhibitor of the amino acid transport in membrane vesicles from *Beta vulgaris*, and, on the other hand, in co-incubation with other protein modifying substances. The calculated reduction is shown in Tables I and/or III. A competition experiment in which the specificity of the transporter could be read off with various amino acids and analogues is shown in Table II for PPP1-20 and in Fig. 4 for AAP2. An analogous experiment in which a competition for histidine uptake in the line JT16::AAP2 was tested is described in Example 5. The time period for PPP1-20 is shown in Fig. 2.

Example 4

Transformation of plants with a construct for overexpression of the coding region of amino acid transporters

From the plasmid pPPP1-20 that contains the cDNA for the amino acid transporter from *Arabidopsis*, an internal fragment of the insert was isolated after BamHI cleavage and cloned in the BamHI cleavage position from pAJ that was first linearized with the enzyme BamHI. Then the cDNA was prepared as the EcoRI/HindIII fragment from pA7 and cloned in the vector pBIN-HYG. After transformation by *Agrobacteria*, this was inserted for infection of leaf segments of tobacco and potato.

Ten independently obtained transformands in which the presence of the intact non-rearranged chimeric gene was demonstrated using Southern blot analysis were tested for modifications of amino acid and nitrogen content. Besides this, amino acid synthesis, photosynthesis rate and transportation were tested.

Example 5

Studies in the uptake of ^{14}C-labeled histidine in the yeast line AAP2

The yeast line JT16::AAP2, obtained in a similar manner to Example 1, was grown in liquid medium until the culture reached the logarithmic phase. After centrifuging the culture, the cells were washed and taken up in 10 mm tris/HCl pH 4.5, 2 mm $MgCl_2$ and 0.6M sorbitol. Around 100 ml of the suspension was added to a solution of 0.5 mm L-histidine plus 1 μCi ^{14}C-labeled L-histidine in 100 μL of the same buffer. The uptake of

the labeled amino acid was measured according to the method described by von Cirillo (1989, Meth Enzymol 174: 617-622). The uptake of the labeled amino acid was compared in a competition experiment with that from different amino acids and analogues in tenfold excess. The relationships are shown in Fig. 5.

Table I

Inhibition of the amino acid transport in 22574d::PPP1-20 - yeast strains by protein modifying substances

	% of transport without inhibitor
0.1 mM DEPC (diethyl pyrocarbonate)	65
10 µM CCCP (Carbonyl cyanide m-chlorophenylhydrazone)	<3
10 µM 2, 4 DNP (Dinitrophenol)	<3
1 mM sodium arsenate	35
10 µM antimycin A	29
500 µM PCMBS (p-chloromercuribenzenesulfonic acid)	78

Table II
Competition by one, fourfold and tenfold excess of amino acids and analogues in 22574d::PPP1-20 - yeast strain

Excess % remaining transport activity:	1 x	4 x	10 x
glutamic acid	64	27	30
aspartic acid	78		27
lysine	86		83
histidine	81	79	58
arginine	85	88	74
threonine	-	50	-
L-proline	49	21	14
D-proline	98		95
3,4-di-OH proline	86		49
azetidine-2-carboxylic acid	91		48
OH-proline	81		45
valine	-	77	47
isoleucine	-	67	-
asparagine	64		57
glutamine	-	27	-
serine	53		18
cysteine	-	21	-
methionine	28		8
glycine	69		16
alanine	55	29	23
leucine	-		-
tyrosine	-		-
tryptophan	82	71	48
phenylalanine	45		16
citrulline		44	
gamma-aminobutyric acid		90	

Table III

Inhibition of the amino acid transports in JT16::AAP2 - yeast strain by protein modifying substances

	% of transport without inhibitor
1 mM DEPC (Diethyl pyrocarbonate)	3.1 ± 1.6
10 μM CCCP (Carbonyl cyanide m-chlorophenylhydrazone)	15.6 ± 2.1
10 μM 2,4 DNP (Dinitrophenol)	7.6 ± 1.6

Although the present invention has been described in relation to particular embodiments thereof, many other variations and modifications and other uses will become apparent to those skilled in the art. Therefore, the present invention is to be limited not by the specific disclosure herein, but only by the appended claims.